

## Informative Identities in the *Begriffsschrift* and ‘On Sense and Reference’

This paper is about the relationship between Frege’s discussions of informative identity statements in the *Begriffsschrift* and ‘On Sense and Reference’. The question of how these discussions relate to one another has a more-or-less standard answer which goes like this. In the *Begriffsschrift* Frege proposes a metalinguistic solution to the puzzle about how an identify statement can be informative. He says that what you find out when you discover that, for example, Hesperus is identical with Phosphorus, is that the two names ‘Hesperus’ and ‘Phosphorus’ are names for the same thing. In ‘On Sense and Reference’ Frege rejects this solution on the ground that it treats identity statements as statements about names rather than statements about objects. His new solution is that if ‘ $a = b$ ’ is potentially informative for co-referring ‘ $a$ ’ and ‘ $b$ ’ this is because ‘ $a$ ’ and ‘ $b$ ’, though they refer to the same object, are associated with different ways of being presented with the object: finding out that  $a = b$  is finding out that the objects presented in these ways are the same.<sup>1</sup>

In this paper I shall argue that, though the standard view of the relationship between the two discussions of informative identity is right in all its skeletal features, there is an important interpretive difficulty that it overlooks. And I shall suggest that

---

<sup>1</sup> See, for example, Michael Dummett *Frege: Philosophy of Language* (London: Duckworth, 1973) (hereafter ‘FPL’), 279. In Michael Thau and Ben Caplan ‘What’s Puzzling Gottlob Frege’ *Canadian Journal of Philosophy* 31 (2001): 159-200, Thau and Caplan argue that Frege never abandoned the *Begriffsschrift* view. I take Heck’s response to their paper (Richard Heck ‘Frege on Identity and Identity Statements: A Reply to Thau and Caplan’ *Canadian Journal of Philosophy* 33 (2003): 83-102) (hereafter ‘Heck “Frege on Identity”’) to provide a decisive refutation of their interpretation. See Heck 83, 101 for re-statements of the standard view. This paper is intended as a continuation of the Thau/Caplan – Heck discussion in that it sets out and addresses a problem for the standard interpretation that Heck’s paper leaves unconsidered.

clearing up this difficulty sheds considerable light on how the explanatory role of Fregean senses should be understood.

The paper has three parts. The first summarises the two discussions of informative identities and sets out the generally overlooked problem raised by Frege's own account of the transition from the *Begriffsschrift* view to the 'On Sense and Reference' view. The second proposes a solution to this problem. The third relates the problem and solution to wider questions about the explanatory role of the notion of the sense of a name.

## **§1 The *Begriffsschrift* and 'On Sense and Reference' treatments of informative identities**

At the start of 'On Sense and Reference' Frege gives a reconstruction of the argument for the *Begriffsschrift* view of identity statements which goes like this:

- i) 'a=a' and 'a=b' can differ in cognitive value even if  $a = b$ . ['a = a' is known *a priori* and cannot extend knowledge; 'a = b' may not be knowable *a priori* and may contain a 'valuable extension of our knowledge']
- ii) An identity statement states either a relation between the objects that the names flanking the identity sign stand for, or a relation between the names themselves.

---

<sup>1</sup> Gottlob Frege, 'On Sense and Meaning', trans. Max Black in Max Black and Peter Geach, eds., *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Blackwell, 1952) (hereafter 'Sense and Reference'), 56.

<sup>2</sup> Frege says '...  $a = a$  holds *a priori* and, according to Kant, is to be labeled analytic....'. For grounds for reading the second of these conditions as 'is incapable of extending our knowledge' see Gottlob Frege, *The Foundations of Arithmetic*, trans. J.L. Austin. (Oxford: Blackwell, 1980) (hereafter '*Grundlagen*'), §88.

iii) If an identity statement states a relation between the objects that the names flanking the identity sign stand for, 'a=a' and 'a=b' cannot differ in cognitive value if  $a = b$ . [This is because if 'a=a' and 'a=b' state relations of this kind and 'a' and 'b' co-refer, each statement states the same relation between a thing and itself.]

therefore

iv) An identity statement states a relation between the names flanking the identity sign.

Here is the *Begriffsschrift* statement of this argument's conclusion:

Equality of content differs from conditionality and negation by relating to names, not to contents.

Elsewhere, signs are mere proxies for their content, and thus any phrase they occur in just expresses a relation between their various contents; but names at once appear *in propria persona* [as themselves] so soon as they are joined together by the symbol for equality of content, for this signifies the circumstance of two names' having the same content. [*Begriffsschrift* sect. 8]

So the suggestion is that 'a=b' is to be treated as short for ' "a" and "b" stand for the same object'. The possible difference in cognitive significance between 'a = a' and 'a = b' is to be explained in terms of the fact that the first sentence says ' "a" stands for what "a" stands for' (or ' "a" and "a" are intersubstitutable') while the second says ' "a" stands for what "b" stands for' (or ' "a" and "b" are intersubstitutable'). For it is never informative to be told that if you substitute a name for itself in a sentence you leave the

---

<sup>1</sup> Gottlob Frege, *Begriffsschrift*, trans. Stefan Bauer-Mengelberg in Jean van Heijenoort (ed.) *From Frege to Godel: A Source Book in Mathematical Logic* (Lincoln, Nebraska: iUniverse.com, 2000) (Hereafter '*Begriffsschrift*').

truth value of the sentence unchanged<sup>5</sup>. But it may be informative to be told that a sentence's truth value remains invariant under replacement of one name by another.

I shall call this account of the potential informativeness of identity statements 'the metalinguistic view'.

In the next part of 'On Sense and Reference' Frege rejects the metalinguistic view. According to his new solution to the puzzle about informative identities, a name is associated with a way of being presented with an object. This way of being presented with an object is the name's 'sense'. The thought expressed by a sentence is the sentence's sense, and is built up out of the senses of the expressions that the sentence contains. It follows that the thoughts expressed by ' $a = a$ ' and ' $a = b$ ' will be different if and only if ' $a$ ' and ' $b$ ' are associated with different ways of being presented with an object, which is to say, if and only if ' $a$ ' and ' $b$ ' differ in sense. Since there can be different ways of being presented with the same object, the thoughts expressed by ' $a = a$ ' and ' $a = b$ ' may differ even though ' $a$ ' and ' $b$ ' co-refer. So ' $a = b$ ' may be informative even though ' $a = a$ ' cannot be.

I shall call this account of the potential informativeness of identity statements 'the sense view'.

The interpretative problem that I want to discuss concerns the passage in which Frege explains his reason for the move to the sense view:

Nobody can be forbidden to use any arbitrarily producible event or object as a sign for something. In that case [and assuming the *Begriffsschrift* view], the sentence  $a = b$  would no longer refer to the subject matter,

---

<sup>5</sup> Here and throughout I follow Frege in assuming that tokens of the same name occurring in a single context share their content. Compare 'Sense and Reference' 58.

but only to its mode of designation; we would express no proper knowledge by its means. If the sign ' $a$ ' is distinguished from the sign ' $b$ ' only as an object (here, by means of its shape) not as a sign (i.e. not by the manner in which it designates something), the cognitive value of  $a = a$  becomes essentially equal to that of  $a = b$ , provided  $a = b$  is true. A difference can arise only if the difference in the signs corresponds to a difference in the mode of presentation of the thing designated.... It is natural, now, to think of there being connected with a sign (name, combination of words, written mark), besides that which the sign designates, which may be called the meaning [reference] of the sign also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained. ['On Sense and Reference', 57. Frege's italics. ]

Here is the generally overlooked point of difficulty in the interpretation of this passage that I want to consider. In the *Begriffsschrift*, Frege suggests that the possibility of a difference in cognitive value between ' $a = a$ ' and ' $a = b$ ' is to be explained by the fact that an identity sentence states a relation between the names flanking the identity sign. But in 'On Sense and Reference' he says that the claim that identity sentences are sentences about names entails that if ' $a$ ' and ' $b$ ' co-refer there is no difference in cognitive value between ' $a = a$ ' and ' $a = b$ ': '...the cognitive value of  $a = a$  becomes essentially equal to that of  $a = b$ , provided  $a = b$  is true...'. So Frege's point in 'On Sense and Reference' is not just, as the standard reading has it,<sup>6</sup> that the metalinguistic view is objectionable because it treats identity sentences as sentences about names. Rather, his claim is that the metalinguistic view is not even a contender as a solution to the puzzle

---

<sup>6</sup> David Kaplan 'Words', *Proceedings of the Aristotelian Society Supplementary Volume LXIV* (1990), 118; Robert May 'Frege on Identity Statements' in C. Cecchetto, G. Chierchia and M. T. Guasti, eds., *Semantic Interfaces: Reference, Anaphora and Aspect* (Stanford: CSLI Publications, 2001), 55; Heck 'Frege on Identity' 98-100. Heck extends the standard reading, suggesting that Frege's abandonment of the *Begriffsschrift* view is motivated at least partly by problems associated with whether the view can cope with sentences in which one of the argument places flanking the identity sign is occupied by a variable (87), and partly by the realization that the puzzle about the difference in cognitive value between ' $a=a$ ' and ' $a=b$ ' generalizes to ' $Fa$ ' and ' $Fb$ ' combined with reluctance to extend the metalinguistic view to all cases where substitution of a co-referring term does not preserve cognitive value (100).

about informative identities because it entails that 'a = b', if true, has the same cognitive value as 'a = a'. And this claim is hard to understand.

The claim is hard to understand because, even if it does not provide a satisfactory explanation of how the difference in cognitive value arises, it seems that the metalinguistic view at least provides a candidate explanation for this difference. On the metalinguistic view, 'a = a' is to be read as short for 'The truth value of '...a...' = the truth value of '...a...'', and 'a = b' is to be read as short for 'The truth value of '...a...' = the truth value of '...b...''. It is plausible that it is always uninformative to be told that the truth value of '...a...' = the truth value of '...a...', but it may be informative to be told that the truth value of '...a...' = the truth value of '...b...'. So it is plausible that the metalinguistic view does generate the possibility of a difference in cognitive value between 'a = a' and 'a = b' where 'a' and 'b' co-refer. It looks like someone objecting to the metalinguistic view must do so on the ground that, though it generates a difference in cognitive value between 'a = a' and 'a = b' it does so in the wrong way. But in the passage from 'On Sense and Reference', Frege is claiming that the metalinguistic view cannot generate the difference at all. And this claim just looks wrong.

This is the generally overlooked problem about the transition from the *Begriffsschrift* to 'On Sense and Reference' that I want to discuss. The rest of the paper is about how the problem is to be solved.

## §2 The rejection of the *Begriffsschrift* view and the explanatory role of Fregean sense

Here is the solution to the problem about Frege's 'On Sense and Reference' attitude to the metalinguistic view that I want to propose. Though Frege writes as if the puzzle about informative identities raised at the beginning of 'On Sense and Reference' and the puzzle discussed in the *Begriffsschrift* are the same, he is in fact concerned with two different puzzles. The metalinguistic view provides a possible solution to the first puzzle. It does not provide a possible solution to the second. Frege's move from the metalinguistic view to the sense view is motivated by the realisation that it is the second puzzle that really matters.

I shall set out the difference between the two puzzles then discuss the difference between the resources required to solve them.

The difference between Frege's two puzzles about informative identities is a difference between senses of 'informative'. In the *Begriffsschrift*, he is concerned with the possibility of what I shall call 'evolutionary informativeness.' The *Begriffsschrift* puzzle about informative identities is a puzzle about how a specific kind of situation and a specific kind of transition in the evolution of the epistemic life of an individual are possible. The situation is the situation of a subject who, without rational inconsistency, understands two co-referring names without realizing that they are names for the same object. The transition is the transition from this situation to the situation in which the subject knows that the names co-refer. So the *Begriffsschrift* puzzle can be posed in terms of the following question:

THE QUESTION ABOUT EVOLUTIONARILY INFORMATIVE IDENTITIES – What account of what understanding a name involves must we give in order to explain the fact that a subject may, without rational inconsistency, understand two co-referential names without knowing that they co-refer, and may then acquire knowledge that the names are in fact names for the same thing?

Put another way, the suggestion is that the difference in cognitive value that Frege is concerned with in the *Begriffsschrift* is a difference in evolutionary cognitive value. If you understand both ‘a’ and ‘b’ then even if you are ideally rational, you might not realise that  $a = b$ : finding out that ‘ $a = b$ ’ is true might constitute an advance in the evolution of your epistemic life. This can never happen with ‘ $a=a$ ’.

In ‘On Sense and Reference’, Frege is no longer concerned with evolutionary informativeness and differences in evolutionary cognitive value. He is concerned with the possibility of what I shall call ‘rational informativeness’. This type of informativeness is bound up with Frege’s conception of how a deductive proof can provide justification for moving from affirmation of its premisses to affirmation of its conclusion. Throughout his career<sup>7</sup>, Frege’s view on this point was that, where a deductive argument justifies moving from a set of premisses to a conclusion, this justification rests on the fact that the argument either comprises or abbreviates a proof whose individual steps are taken in accordance with basic laws of inference. He thought that a proof laid out as a series of

---

<sup>7</sup> For illustrative passages see *Begriffsschrift*, preface 5-6; *Grundlagen* introduction p. IX, 102-103 (§§90-91); Gottlob Frege, *Grundgesetze der Arithmetik* vol. 1, excerpted in Black and Geach (1952) (hereafter ‘*Grundgesetze*’), 118-119, 128.



steps of this kind displays the ‘ultimate ground’<sup>8</sup> for moving from its initial premisses to its conclusion: it shows how the move from premisses to conclusion is in keeping with ‘the connexion and natural order of truths’<sup>9</sup>. A proof of this kind is built up out of ‘logically’, as opposed to just ‘intuitively’<sup>10</sup> self-evident steps. And if this kind of proof has premisses which are themselves purely logical, it reveals the logical self-evidence of its conclusion.

Frege’s picture of proofs as built up out of logically self-evident steps, and of a rational order of logically self-evident relations between thoughts, carries with it the possibility of a distinctive kind of difference in cognitive value between extensionally equivalent expressions. This is the possibility that two expressions might stand for the same object (if they are names) or the same function (if they are predicates) but that a chain of inference constructed using one expression might be logically self-evident while the parallel chain of inference constructed using the other is not. For example, consider the following proof in Peano Arithmetic (recall that ‘S0’ says ‘the successor of 0’, ‘SS0’ says ‘the successor of the successor of zero’ (so ‘SS0’ is a name for 2), and so on, so that what the proof shows is that  $1 + 2 = 3$ . The axioms in lines 1 and 2 give the Peano Arithmetic characterisation of ‘+’):

1  $\forall x(x + 0 = x)$  (Axiom)

2  $\forall x \forall y(x + Sy = S(x + y))$  (Axiom)

3  $S0 + 0 = S0$  (from 1 by Universal Instantiation)

---

<sup>8</sup> *Grundlagen* §3 p.3.

<sup>9</sup> *Grundlagen* §17 p.23.

<sup>10</sup> See *Grundlagen* §93 p. 102 for the distinction between ‘logical’ and ‘intuitive’ self-evidence, and see *Grundlagen* §16 p. 23, §80 p. 93 for the same distinction put in different terms.

4  $\forall y(S0 + Sy = S(S0 + y))$  (from 2 by Universal Instantiation)

5  $S0 + S0 = S(S0 + 0)$  (from 4 by Universal Instantiation)

6  $S0 + S0 = SS0$  (from 3, 5 by Substitution)

7  $S0 + SS0 = S(S0 + S0)$  (from 4 by Universal Instantiation)

8  $S0 + SS0 = SSS0$  (from 6, 7 by Substitution)

If we allow that transitions taken in accordance with the rules of inference the proof employs (Substitution and Universal Instantiation) are logically self evident, 1-8 constitute a series of logically self-evident steps from 1 and 2 to the conclusion, 8. But now consider what happens when we replace 'SS0' in all its occurrences with the co-referential ' $S0 \times SS0$ ':

1  $\forall x(x + 0 = x)$  (Axiom)

2  $\forall x \forall y(x + Sy = S(x + y))$  (Axiom)

3  $S0 + 0 = S0$  (from 1 by Universal Instantiation)

4  $\forall y(S0 + Sy = S(S0 + y))$  (from 2 by Universal Instantiation)

5  $S0 + S0 = S(S0 + 0)$  (from 4 by Universal Instantiation)

6\*  $S0 + S0 = S0 \times SS0$  (the result of replacing 'SS0' in 6 with ' $S0 \times SS0$ ')

7\*  $S0 + (S0 \times SS0) = S(S0 + S0)$  (the result of replacing 'SS0' in 7 with ' $S0 \times SS0$ ')

8\*  $S0 + (S0 \times SS0) = S(S0 \times SS0)$  (the result of replacing 'SS0' in 8 with ' $S0 \times SS0$ ')

The resulting series of statements no longer constitutes a proof in Peano Arithmetic. 6\*, 7\*, and 8\* are all true. But 6\* and 7\* no longer follow by logically self-evident steps from the lines above them. In the initial proof, 6 was got from 3 and 5 by a logically self-evident step (an application of Substitution). But Substitution does not carry us from 3 and 5 to 6\*. Similarly, 7 in the initial proof was got from 4 by a logically self-evident step (an application of Universal Instantiation). But Universal Instantiation does not take us from 4 to 7\*. So the replacement of 'SSO' with the co-referring 'S0 × SS0' has turned a proof into a non-proof. Substitution of co-referring expressions preserves truth. But it does not preserve the location of the statement in which the substitution occurs in the network of logically self-evident rational relations between statements.

I shall call this kind of difference in cognitive value 'rational distance'. In general, the notion of 'rational distance' between names can be defined like this<sup>11</sup> (where a chain of inference is 'logically self evident' if and only if each of its steps is):

There is 'rational distance' between names 'a' and 'b' if and only if, for some  $\gamma_1, \dots, \gamma_n$ , where  $\gamma_1, \dots, \gamma_n$  is a chain of inference which contains occurrences of 'a' but not of 'b'<sup>12</sup>, replacing all occurrences of 'a' with occurrences of 'b' transforms  $\gamma_1, \dots, \gamma_n$  from a chain of inference which is logically self-evident to one which is not, or from a chain of inference which is not logically self-evident into one which is.

---

<sup>11</sup> Compare *Begriffsschrift*, 12: '...the contents of two judgments may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgments, always follow also from the second, when it is combined with these same judgments, and conversely, or this is not the case....Now I call that part of the content that is the *same* in both the *conceptual content*.' [Frege's italics.]

<sup>12</sup> The stipulation that 'b' does not occur in  $\gamma_1, \dots, \gamma_n$  is required to set aside cases like the transformation of 'Fa, Gb, therefore  $\exists x(Fx \ \& \ Gx)$ ' into 'Fb, Gb, therefore  $\exists x(Fx \ \& \ Gx)$ ' and ' $\sim Fa$ , Fb, therefore p' into ' $\sim Fb$ , Fb, therefore p'.

Given this account of rational distance between names, 'rational informativeness' for identity statements can be defined like this:

An identity statement between co-referring names 'a' and 'b' is 'rationally informative' if and only if there is rational distance between 'a' and 'b'.

The 'On Sense and Reference' question about informative identities is the question of how rational informativeness is possible:

THE QUESTION ABOUT RATIONALLY INFORMATIVE IDENTITIES – What account of what a subject's understanding of a name involves must we give in order to allow for the fact that there might be rational distance between 'a' and 'b' even though 'a' and 'b' co-refer?

Now I am going to argue that different resources are required to answer the question about evolutionarily informative identities and the question about rationally informative identities: the question about evolutionarily informative identities can be answered using only the resources Frege allows himself in the *Begriffsschrift*, but the question about rationally informative identities requires something more.

In the *Begriffsschrift* Frege recognizes only two kinds of content which might determine an expression's place in the natural order of thoughts. Firstly, there is ordinary content: a name's ordinary content is an object; a predicate's ordinary content is a function. Secondly, there is metalinguistic content: names occurring in identity

statements are taken to represent themselves rather than their ordinary contents. So the proposal is, in effect, that names are labels on objects and predicates are labels on functions, and the roles that names and predicates play are to be explained either in terms of the things they stand for or in terms of the fact that they are labels for these things. This proposal can answer the question about evolutionarily informative identities. For there is nothing incoherent about the possibility that a single object may have two labels and a subject may use both labels without realising that they are labels on the same object.

But the *Begriffsschrift* view does not have the resources to answer the question about rational informativeness: it cannot explain how there might be rational distance between co-referring names. To see why not, consider again the contrast between the Peano Arithmetic proof and the non-proof got by replacing all instances of ‘SS0’ with ‘S0 × SS0’ which illustrates the rational distance between these two expressions. What is required to explain how this rational distance arises is an account of why ‘SS0’ and ‘S0 × SS0’ should interact differently with the axioms and rules of inference used to construct the proof. The expressions stand for the same object, so the difference cannot be explained by appeal to a difference at the level of ordinary content (reference). So if we are to explain the fact that ‘SS0’ and ‘S0 × SS0’ interact differently with the axioms and rules of inference using only the resources of the *Begriffsschrift*, we must do so in terms of a difference in metalinguistic content – in terms of the difference between the signs ‘SS0’ and ‘S0 × SS0’ themselves. But the mere shape of a sign cannot, in general, determine which series of sentences containing the sign will count as a proof. For we are

free to use any object as a sign for any other object: ‘Nobody can be forbidden to use any arbitrarily producible event or object as a sign for something’<sup>13</sup>. And because our choice of which signs to use for which things is arbitrary in this way, differences between signs do not, in general, map onto differences between ways in which signs may be used to construct logically self-evident proofs.

I suggest that the primary explanatory role of Fregean sense<sup>14</sup> is to explain the possibility of rational distance between extensionally equivalent expressions. From the time of the discovery of the distinction between sense and reference onwards, Frege thinks that co-referring names may fit differently into the ‘natural order’ of thoughts because a name’s position in this order is determined by the way of being presented with an object with which it is associated, and the same object may be presented in different ways.<sup>15</sup>

So here is the solution to the problem about Frege’s transition from the metalinguistic view to the sense view that I want to propose. The metalinguistic view provides a coherent possible response to the question about evolutionary informativeness. But it does not provide a coherent possible answer to the question about rational informativeness. In the first part of the passage from ‘On Sense and Reference’ that I have quoted Frege is telling us why not. His argument, with some details filled in, is as follows:

---

<sup>13</sup> This is the first sentence of the quotation from ‘Sense and Reference’ given on p. 2.

<sup>14</sup> The secondary roles are to be the oblique referent of an expression occurring in a ‘that’ context (‘Sense and Reference’, 58-59, 65) and to be what is passed from speaker to hearer in successful communication (‘Sense and Reference’, 59).

<sup>15</sup> See *Grundgesetze* vol. 1 §5 for explicit statement of the claim that steps in a proof are steps between thoughts.

1 'a = b' is shorthand for '“a” and “b” can be substituted for one another without changing the truth values of sentences in which the substitutions occur'. [The metalinguistic view assumed for *reductio*.]

2 To add 'a = b' to a chain of inference is to add only the information that 'a' and 'b' can be substituted for one another without changing the truth value of the sentence in which the substitution occurs. [From 1]

3 Let  $\gamma_1, \dots, \gamma_n$  be a chain of inference containing 'a' but not 'b'. The addition of 'a = b' entitles us to extend this chain of inference as follows:

$\gamma_1, \dots, \gamma_n, a = b, \gamma_{n+1}, \dots, \gamma_{n+n}$

where  $\gamma_{n+1}, \dots, \gamma_{n+n}$  is the chain of inference got from  $\gamma_1, \dots, \gamma_n$  by replacing 'a' in all its occurrences with 'b'.

4 Now suppose that  $\gamma_{n+1}, \dots, \gamma_{n+n}$  is logically self-evident while  $\gamma_1, \dots, \gamma_n$  is not. This is to suppose that whatever information 'a = b' carries enables us to transform a chain of inference which is not logically self evident into one which is. So it is to suppose that the information that 'a = b' brings to the proof is information above and beyond just the information that 'a' can be substituted for 'b' in all its occurrences without changing the truth value of the sentence in which the substitution occurs. This contradicts 2.

So

5 If  $a = b$  and  $\gamma_1, \dots, \gamma_n$  is a chain of inference containing 'a' but not 'b', the chain of inference got from  $\gamma_1, \dots, \gamma_n$  by replacing every occurrence of 'a' with an occurrence of 'b' is logically self evident if and only if  $\gamma_1, \dots, \gamma_n$  is.

But

6 There are many counterexamples to 5 (many instances in which ' $a = b$ ' 'expresses proper knowledge').

so

7 The metalinguistic view must be rejected.

This solution to the problem set out in §1 also solves a more widely acknowledged interpretive problem about the relationship between the accounts of informative identities in the *Begriffsschrift* and 'On Sense and Reference'. This is the problem of how to understand Frege's appeal to ways of being presented with objects in the *Begriffsschrift*.<sup>16</sup> In the problem passage, Frege is considering a possible objection to the metalinguistic view of identity statements. This is the objection that the metalinguistic

---

<sup>16</sup> Thau and Caplan base their claim that Frege never rejected the *Begriffsschrift* view partly on this problem: 'What's Puzzling Gottlob Frege', 177-178.



view leaves us with no account of why we need the identity sign: if 'a=b' says just that 'a' and 'b' are intersubstitutable, then so long as we can avoid having two signs for the same object we will have no need for an identity sign. In reply to this objection, Frege sets out an example in which two names for a single geometrical point correspond to two ways of determining or describing it. And he writes

To each of these two ways of determining [the point] there answers a separate name. The need for a symbol of equality of content thus rests on the following fact: The same content can be fully determined in different ways; and *that*, in a particular case, *the same* content actually is given by *two ways of determining it*, is the content of a *judgment*. [*Begriffsschrift* §8, Frege's italics]

This passage raises a question which must be answered by any account of the move from the metalinguistic view of identity statements to the sense view: How does the role played by 'ways of determining' in the *Begriffsschrift* differ from the role played by 'modes of presentation' in 'On Sense and Reference'?

The interpretation of the move from the metalinguistic view to the sense view that I have presented suggests the following answer to this question. In the *Begriffsschrift*, the role of 'ways of determining' is to explain why the possibility of evolutionary informativeness will always arise, and so why we need the identity sign. Frege's point is that the possibility of evolutionary informativeness cannot be avoided because we cannot avoid determining or describing the same object in different ways. In our thinking about particular things, we gather together the beliefs acquired using a single way of determining or thinking about an object as beliefs to be expressed using a single name. But it will always be possible to determine the same object in different ways. So it will

always be possible to employ different names for the same object. And we will always need the identity sign to express what is discovered when we find out that two names co-refer.

So in the *Begriffsschrift* Frege does assign an explanatory role to different modes of presentation for the same object. But this is a much smaller role than the role he assigns them in 'On Sense and Reference'. In the *Begriffsschrift* the patterns of rational relations laid down by proofs are made up only of expressions and their ordinary contents (their referents) and Frege appeals to modes of presentation to explain why a specific kind of pattern will always arise. By the time of 'On Sense and Reference' he thinks that modes of presentation are the stuff from which the pattern itself is made.

### **§3 Consequences**

The question of how to understand the transition from the *Begriffsschrift* account of informative identities to the 'On Sense and Reference' account is historically important in its own right. But I take it that the main non-historical reason to try to get straight about this question is to see what a right account of the transition can tell us about the explanatory role of the notion of sense. I shall close with some preliminary remarks on this issue.

Consider a standard formulation of the main criterion that philosophers in the Fregean tradition have treated as determining whether two names for an object share a sense:

THE INTUITIVE CRITERION OF DIFFERENCE FOR SENSES – ‘a’ and ‘b’ differ in sense if and only if it is possible for a subject who understands ‘a’, ‘b’, and ‘F’ to affirm ‘Fa’ and deny ‘Fb’ without loss of rational coherence.<sup>17</sup>

Operating with this criterion, philosophers have concluded (i) that every instance in which it is possible for a subject who understands ‘a’, ‘b’, and ‘F’ to affirm ‘Fa’ and deny ‘Fb’ without loss of rational coherence is to be explained by appeal to a difference in sense between ‘a’ and ‘b’, and (ii) that where ‘a’ and ‘b’ share a sense a rational subject who understands ‘a = b’ must accept it as true. But, if what I have argued in this paper is right, the intuitive criterion of difference needs revision, and both of the conclusions standardly drawn from it are wrong.

The intuitive criterion of difference needs revision because it ignores the distinction between evolutionary and rational informativeness. If the argument of this paper is right, what is required is a criterion for difference in sense which aligns difference in sense with rational distance. So we should say something like

THE REVISED CRITERION OF DIFFERENCE FOR SENSES – ‘a’ and ‘b’ differ in sense if and only if, for some  $\gamma_1, \dots, \gamma_n$ , where  $\gamma_1, \dots, \gamma_n$  is a chain of inference which contains occurrences of ‘a’ but not ‘b’, replacing all occurrences of ‘a’ with occurrences of ‘b’ transforms  $\gamma_1, \dots, \gamma_n$  from a chain of inference which is logically self-evident to one which is not, or from a chain of inference which is not logically self-evident to one which is.

---

<sup>17</sup> This formulation is derived from Evans’s ‘intuitive criterion of difference’ for the senses of sentences – see Gareth Evans *The Varieties of Reference* (Oxford: Oxford University Press, 1982), 19–20. 20. See Richard Heck ‘Do Demonstratives have Senses?’, *Philosophers’ Imprint* 2 (2002), 11, 21 for an example of this kind of criterion at work.

Given this new criterion for sameness of sense, it no longer follows that every situation in which a subject can take different propositional attitudes towards 'Fa' and 'Fb' without loss of rational coherence is to be explained by appeal to a difference in sense. Here is an example where, given the revised criterion of difference for senses, this kind of explanation is not appropriate. Suppose that you are reading a variety of sources on philosophical, social, and political activity in early Twentieth Century England. You keep finding references on the one hand to someone known as 'Bertie', and on the other to someone known as 'Russell'. You do not realize that the two names co-refer. In this case it seems fair to say that you understand both 'Bertie' and 'Russell'. It also seems fair to say that you would be making a factual mistake, rather than a mistake of rationality, if you affirmed 'Bertie was a witty correspondent' while denying 'Russell was a witty correspondent' – in affirming the first statement while denying the second you would suffer no loss of rational coherence. And it is perfectly possible for this situation to arise even though you do not associate modes of presentation with 'Bertie' and 'Russell' which endow the names with different inferential properties: a chain of inference,  $\gamma_1, \dots, \gamma_n$ , containing 'Bertie' is logically self-evident if and only if the parallel chain of inference got by replacing 'Bertie' with 'Russell' in  $\gamma_1, \dots, \gamma_n$  is. In this case the differences between the beliefs you would express using 'Bertie' and those you would express using 'Russell' are not generated by a difference between the logically self-evident inferences containing 'Bertie' and the logically self-evident inferences containing 'Russell'. So they are not generated by a difference in sense. They are generated by a difference between the

information expressed using ‘Bertie’ and the information expressed using ‘Russell’ that happens to have come your way.

Though it is not possible to argue for this claim here, I suggest that all cases of informative identity involving ordinary proper names are cases of this kind. Descriptions (like ‘SSO’ and ‘S0  $\times$  SSO’) are associated with modes of presentation which may endow them with distinctive inferential properties. There is a hard question about whether demonstratives and indexicals are associated with modes of presentation which might have this effect too. But there are good arguments for the conclusion that an ordinary proper name is not associated with any distinctive way of identifying its bearer, except as the bearer of the name.<sup>18</sup> In this case the inferential properties of ordinary proper names – which chains of inference containing them are logically self evident – are determined just by their bearers and their status as labels for their bearers. So, the potential informativeness of ‘a = b’ where ‘a’ and ‘b’ are ordinary proper names is mere evolutionary informativeness, not to be explained by appeal to a difference in sense.

The revised criterion of difference for senses also makes room for a more delicate kind of case of evolutionary informativeness without difference in sense. Examples of this more delicate kind can be extracted from Frege’s own work on definitions. At *Grundlagen* §68 Frege proposes the following definition of the concept of direction:

The direction of line *a* is the extension of the concept ‘parallel to line *a*’. (1)

---

<sup>18</sup> See for example Scott Soames, *Beyond Rigidity* (Oxford: Oxford University Press, 2002) especially ch. 2 and pp. 63–64. Soames is expanding on the arguments against description theories of names in Saul Kripke, *Naming and Necessity* (Oxford: Blackwell, 1972), Lecture 2.

Frege acknowledges that the correctness of this definition is not, at first, intuitively transparent.<sup>19</sup> This is to acknowledge that, where L is a name for a line, the identity statement

The direction of L = the extension of the concept 'parallel to L'. (2)

may come as a surprise to an otherwise competent speaker. So Frege needs an account of what you learn when you realise that the definition is correct and the identity statement true.

Neither the metalinguistic view of informative identities nor the reading of the sense view embodied by the intuitive criterion of difference for senses can provide a satisfactory answer to this question. I shall set out why not, then say how the account of the explanatory role of the notion of sense suggested in this paper does better.

According to the metalinguistic view, the statement 'The direction of L = the extension of the concept "parallel to L"' is to be read as short for 'The names "the direction of L" and "the extension of the concept 'parallel to L'" refer to the same object'. This account secures the evolutionary informativeness of the identity statement. But it leaves us without an account of what is special about statements of definitional equivalence. According to the metalinguistic view, when you discover the truth of (2) you make the same kind of cognitive advance as you make when you discover the truth of

---

<sup>19</sup> See his comment on the analogous definition for 'the number of F's' at the start of §69: 'That this definition is correct will perhaps be hardly evident at first'.

The direction of line L = the direction in which I am now walking. (3)

The metalinguistic view cannot furnish an account of the distinctive role of statements of definitional equivalence in enabling us to construct right arguments concerning whatever is defined.

If we stay with the reading of the move to the sense view embodied by the intuitive criterion of difference for senses, it is hard to see how the discovery of the distinction between sense and reference improves this situation. For according to this reading the move to the sense view leaves Frege with a choice between two accounts of (2). On the first account, the names flanking the identity sign differ in sense, and the statement is potentially informative. On the second, the expressions have the same sense, so the identity statement cannot be doubted by anyone who understands it without loss of rational coherence. The first option is unsatisfactory because, like the metalinguistic view, it leaves us without an account of the significance of statements of definitional equivalence. If the informativeness of (2) is to be explained just by saying that ‘the direction of L’ and ‘the extension of the concept “parallel to L”’ differ in sense, we are explaining the informativeness of (2) in the same way as we explain the informativeness of (3), and are left without an account of the sense in which the definition should be regarded as ‘giving the meaning’ of the defined term. The second option entails that if you doubt (2) then either you do not understand all of the terms it contains or you are guilty of rational inconsistency. This option is unsatisfactory because it leaves us with no account of how progress towards discovering the correctness of the definition is possible. In order to progress towards the definition you must understand the terms involved. But if

you understand the terms involved yet doubt the definition you are, on this view, rationally incoherent in your thinking involving them. In discovering the truth of the definition you somehow progress from this position of rational incoherence to a realisation which sweeps the incoherence away. But it is completely obscure how this step can be made.

The interpretation set out in §2 of this paper suggests a better account of the informativeness of statements of definitional equivalence. This is because it carries with it the possibility of a gap between the inferential properties that may be associated with an expression by a speaker who counts as understanding it, and the deeper inferential properties on which the inferential patterns the speaker associates with the expression rest. An expression's sense determines its place in the pattern of deeper inferential connections: this is Frege's 'natural order' of logically self-evident inferential relations between statements. And all right deductive reasoning owes its justification to this pattern of logically self-evident relations. So the most fundamental justification for the use of an expression is justification in terms of its sense. But not all justification is fundamental justification. The view of the explanatory role of the notion of sense suggested in §2 enables us to say that before recognizing the correctness of the definition you are operating with expressions which you do understand but for whose use you have only non-fundamental justification. Before coming to accept the definition you are able to make some justified moves involving the expressions. These moves are 'justified' in that they do not lead you into contradiction and you may have some inchoate grasp of why they are the right moves to make. But before recognising the definition's correctness you do not have fundamental justifications for the moves you make using the defined term. In



looking for the definition you are looking for the fundamental justification which lies beneath your non-fundamental justification: the position in the network of logically self-evident inferential relations which will explain why the non-fundamental inferences you have been engaged in so far work.

This is only a preliminary account of the implications of the shift in perspective about Fregean sense that I am suggesting. I hope in another paper to explore the implications of the move to the revised criterion of difference for the debate about whether Frege's notion of sense, introduced in terms of its primary explanatory role, can perform the secondary roles he assigns it. Frege wants the sense of 'a is F' (the thought that the sentence expresses) to be both what is passed from speaker to hearer in successful communication, and what 'a is F' contributes to determining the truth value of a sentence in which it occurs in a 'that' context (a sentence of form 'S believes that a is F'<sup>20</sup>). The extant debate about whether Frege's notion of sense can perform all three of these explanatory roles<sup>21</sup> assumes the standard criterion of difference for senses, and asks whether this criterion can also be held to individuate what is shared between speakers in successful communication, and the referents of expressions when they occur in attitude ascribing contexts. If the standard criterion of difference for senses is replaced with the revised criterion, the question of whether the notion of sense can perform all of the roles that Frege assigns it will end up looking very different from the way it looks in the current debate. But I cannot pursue this question further here.

---

<sup>20</sup> See note 14.

<sup>21</sup> For a canonical statement of the anti-Fregean side of this debate see John Perry, 'Frege on Demonstratives' *The Philosophical Review* 86 (1977): 474-497. For a Fregean reply see Gareth Evans, 'Understanding Demonstratives' reprinted in Evans *Collected Papers* (Oxford: Oxford University Press, 1985), 291-321. For a more recent discussion see Richard Heck, 'The Sense of Communication' *Mind* 104 (1995), 79-106.

## §4 Conclusion

I take the moral of the interpretative problem I have discussed to be this. The explanatory role that Frege intends senses to fulfill entails a precise criterion for sameness and difference in sense. Frege intends an expression's sense to be its contribution to determining where the thoughts expressed by sentences containing it lie in the natural order of thoughts: the pattern of thoughts determined by logically self-evident relations between them. So expressions differ in sense if and only if substituting one expression for another in a sentence changes the location of the thought it expresses in this natural order. The traditional focus on the role of Fregean sense in explaining potentially informative identities has skewed this precise criterion into a blunt one. Philosophers have thought that within a Fregean framework potential informativeness must always signal a difference in sense. And they have thought that where 'a' and 'b' share a sense 'a = b' cannot be potentially informative. But to draw these conclusions is to run together evolutionary and rational informativeness and fail to recognise the distinction between what is required to explain them. In terms closer to Frege's own, it is to mistake the patterns of rational relations between our beliefs which arise from the ways we form them for the deeper patterns of rational relations between belief contents which provide the ultimate justification for our inferential practices.<sup>22</sup> My suggestion here has

---

<sup>22</sup> Compare *Grundlagen* §3 p.3.

been that a right account of Frege's move to the 'On Sense and Reference' view of identity statements provides a first step towards setting this mistake right.<sup>23</sup>

---

<sup>23</sup> Thanks to Kevan Edwards, Gurpreet Rattan, David Velleman, and two anonymous referees for this journal for comments on drafts of this paper. Thanks also to audiences at the 2006 congress of the Canadian Philosophical Association, Oxford University, and the University of Canterbury, New Zealand.